Calculation of f-Divergence

Yu Du, Jonathan C. Wong
Institute of Modern Physics, Chinese Academy of Sciences

October 12, 2026

f-Divergences

The f-divergences are a common class of methods used to measure the difference between two distributions, defined as follows:

$$D_f[\rho_1(\mathbf{x}) \parallel \rho_2(\mathbf{x})] = \int \rho_2(\mathbf{x}) f\left[\frac{\rho_1(\mathbf{x})}{\rho_2(\mathbf{x})}\right] \mathrm{d}\mathbf{x},$$

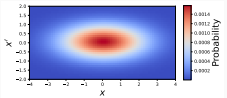
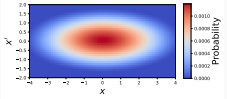
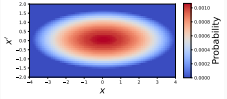
- $f(\cdot)$ is a convex function and satisfies $f(1) = 0$;
- Different $f(\cdot)$ correspond to different statistical divergences:

Name	$f(t)$	h	**Range**
Kullback-Leibler	$t \ln t$	$\rho_1(\mathbf{x}) \ln\left[\frac{\rho_1(\mathbf{x})}{\rho_2(\mathbf{x})}\right]$	$[0, \infty)$
Jensen-Shannon	$\frac{1}{2}\left[\ln\left(\frac{2}{t+1}\right)^{t+1} + t \ln t\right]$	$\frac{1}{2}\left\{\rho_2(\mathbf{x}) \ln\left[\frac{2\rho_2(\mathbf{x})}{\rho_1(\mathbf{x})+\rho_2(\mathbf{x})}\right] + \rho_1(\mathbf{x}) \ln\left[\frac{2\rho_1(\mathbf{x})}{\rho_1(\mathbf{x})+\rho_2(\mathbf{x})}\right]\right\}$	$[0, \ln 2]$
Total Variation	$\frac{1}{2}\lvert t-1\rvert$	$\frac{1}{2}\lvert\rho_1(\mathbf{x}) - \rho_2(\mathbf{x})\rvert$	$[0, 1]$
Squared Hellinger	$\frac{1}{2}(\sqrt{t}-1)^2$	$\frac{1}{2}\left[\sqrt{\rho_1(\mathbf{x})} - \sqrt{\rho_2(\mathbf{x})}\right]^2$	$[0, 1]$

Four forms of f-divergences; $t = \rho_1(\mathbf{x})/\rho_2(\mathbf{x})$; $h = \rho_2(\mathbf{x}) \cdot f[\rho_1(\mathbf{x})/\rho_2(\mathbf{x})]$; $\mathbf{x} \in \mathbb{R}^{2n}$

Distribution Given as a Probability Density Function (PDF)

Several common 4D beam distributions:

Distribution with Elliptical Symmetry	**Definition** $\rho(x, x', y, y') = \rho(I)$ $I = \mathbf{x}^T\Sigma^{-1}\mathbf{x}$	**Schematic Diagram of 2D Projection** (x, x')
Gaussian	$\frac{1}{(\sqrt{2\pi})^4\lvert\Sigma\rvert^{\frac{1}{2}}} \cdot e^{-\frac{1}{2}I}$	
Parabolic	$\frac{6}{(\sqrt{8\pi})^4\lvert\Sigma\rvert^{\frac{1}{2}}} \cdot \left(1 - \frac{I}{8}\right),\ I < 8$	
Waterbag	$\frac{2}{(\sqrt{6\pi})^4\lvert\Sigma\rvert^{\frac{1}{2}}},\ I < 6$	

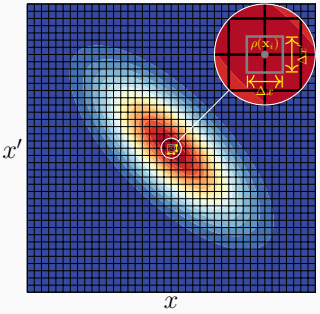
Quadratic form expression of beam distributions with elliptical symmetry; $\mathbf{x} = (x, x', y, y')^T$*;*
Σ *is a covariance matrix composed of 10 independent second-order moments.*

Distribution Given as a Probability Density Function (PDF)

Discretize the continuous phase space into regular grid cells:

$$D_f[\rho_1(\mathbf{x}_i)||\rho_2(\mathbf{x}_i)] = \sum_{i=1}^{m} \rho_2(\mathbf{x}_i) f\left[\frac{\rho_1(\mathbf{x}_i)}{\rho_2(\mathbf{x}_i)}\right] \Delta V$$

- $\mathbf{x}_i = (x_{1i}, x'_{1i}, x_{2i}, x'_{2i})$ is the coordinate of the i-th node.
- $\Delta V = \Delta x_1 \Delta x'_1 \Delta x_2 \Delta x'_2$ is the volume of each grid cell.
- The total number of nodes is $m = \left(\frac{l}{d} + 1\right)^4$, which determines the accuracy. Here l is the range of each axis and d is the cell length.



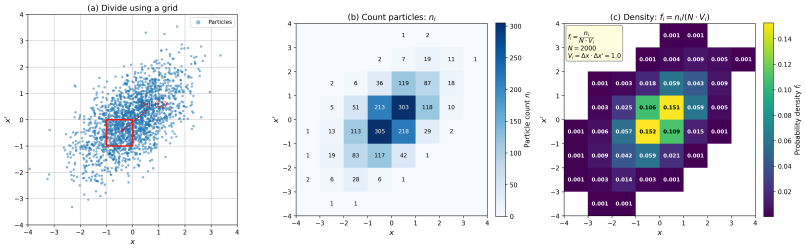
Distribution Given as Particle Coordinates in Phase Space

Given only **particle coordinates** $\{(x, x', y, y')\}$: histogram method.

- Divide the particle distribution using a grid;
- Count the number of particles in each grid cell;
- Calculate the probability density of each grid cell:

$$f_i = \frac{n_i}{N \cdot V_i}$$

where f_i is the probability density of the i-th grid cell, n_i is the count of the i-th grid cell, N is the total count, and V_i is the volume of the i-th grid cell.

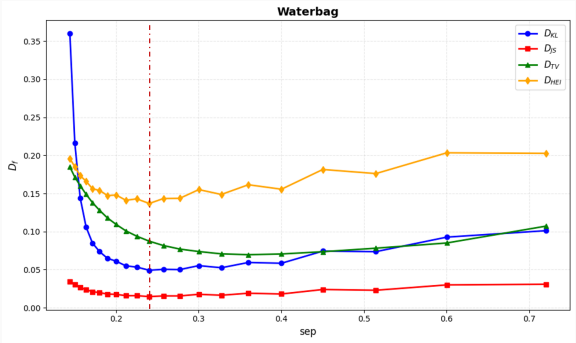


Distribution Given as Particle Coordinates in Phase Space

Important — choice of parameters

Number of particles N and grid resolution Δx must be *consistent*:

- too fine a grid $\Rightarrow$ many empty cells; too coarse $\Rightarrow$ loss of structure.



D_f between converted and true density distributions.

- For the waterbag distribution, if the number of particles is 10^7;
- The corresponding grid resolution $\Delta x = 0.24$ minimizes the conversion error.

Conclusions

Case 1: Distribution given as probability density function (PDF)

- Ideal distributions can be generated directly in **density form**.
- f-divergence is computed directly on the discretized density grid.

Case 2: Distribution given as particle coordinates in phase space

- First convert particle coordinates into a **density form** (e.g., via histogram).
- Caution: choose an appropriate number of particles and grid density.
- Then apply the Case 1 method to compute f-divergence.

Thanks !